

**Minutes**  
**State Board of Education AIMS Special Session**  
**May 11, 2005**

The State Board of Education held an AIMS Special Session on May 11, 2005 at the Hilton Pointe Squaw Peak Resort, 7677 North 16<sup>th</sup> Street, Phoenix, Arizona 85020. The meeting was called to order at 6:30 PM.

Pledge of Allegiance, Moment of Silence and Roll Call

**Members Present**

Dr. Matthew Diethelm, President  
Ms. JoAnne Hilde, Vice President  
Superintendent Tom Horne  
Mr. Jesse Ary  
Ms. Nadine Mathis Basha  
Ms. Joanne Kramer  
Ms. Anita Mendoza  
Dr. Karen Nicodemus  
Ms. Cecilia Owen  
Dr. John Pedicone

**Members Absent**

Dr. Michael Crow

**1. CALL TO THE PUBLIC**

Ms. Charlene Mendoza, AmeriSchools, Tucson, Arizona, noted the following regarding the Standards setting and AIMS:

- The process is good;
- Scores are fair and accurate;
- She has a great deal more faith today regarding the integrity of the process;
- There is opportunity for educators in the classroom to look at the tests without outside influence;
- She appreciated being included in the process and is proud to be a teacher in Arizona;
- Concerns/Questions:
  - Public perception that this is a graduation test and the fact this it has graduation implications for students;
  - AIMS has had a big influence on teaching; writing has improved greatly; AIMS has been effective in the public education system to design and set standards.

**2. Presentation, Discussion and Possible Consideration to Approve Proposed Performance Level Standards for Arizona's Instrument to Measure Standards**

Dr. Karla L. Egan, Senior Research Manager, CTB/McGraw-Hill, stated that the past three days have been very productive in bookmark standard setting for grades 3, 5, 8 and high school reading and writing and body of work for grades 3, 5, 8, and high school writing. Dr. Egan added that the quality of teachers was excellent, the quality of conversations was on target, and that they were effective and efficient workers who should be commended on their efforts for the past three days.

Dr. Egan noted a change from the previous discussions regarding the process:

- There were to be three rounds of voting and discussion followed by the table leaders smoothing the impact data;

- All 3 rounds for bookmark happened yesterday;
- Groups wanted more information from last year's test, looking at percent correct, and the high school graduation rates to help them make informed decisions;
- Met this morning in large groups with the high school group together and reading, writing and math groups for grades 3, 5, and 8 together;
- After this meeting, the groups went to breakout rooms for discussion then voting;
- Grade 8 reading wanted a 5<sup>th</sup> round of voting which was appropriate;
- Table leaders from content areas then began smoothing data

Dr. Egan presented the results (see materials in packet) that represent the impact data and the data smoothed by the table leaders. Based on the smoothing of the data, the percent of meets and above is nearly 67% in grade 3, nearly 68% in grade 5, nearly 69% in grade 8 and about 71% in grade 10. Dr. Egan noted the memo written to the State Board by the evaluators cites the criteria used by the team in smoothing and eliminating dips and bumps between grade levels. The memo also states that the team's "recommended cut scores are fair, reasonable and attainable by Arizona students".

Further discussion ensued around the following points:

- Increased and/or decreased the number of items students are expected to perform on;
- How percentage changes to meet standards in '05 as compared to '04 results;
  - Dr. Egan noted that the tests are better aligned as the cut score previously was not aligned to what was expected of Arizona students. There are more realistic expectations.
- Better job of aligning, better curriculum, better efforts
- With extensive professional development, teachers sense they are better teachers than in the past because there is a better understanding of aligned curriculum
- Teachers are able to teach at a better level

Reading:

- Standard Error of Measurement (SEM) tables are based on recommendations of the entire community and not based on smoothed data
- Recommended cut points are in the middle of the table
- If cut score is lowered by 1 SEM, impact is interesting but skews in favor of "exceeds" rather than "meets"
- Greatest impact seen at the "thicker" part of the curve as there are fewer students at the lower end of the curve
- Policy-based reasons for adjusting cut scores
  - Teachers have a perspective and State Board members may have another viewpoint
  - Movement from classification to classification because of the SEM is somewhat dependent on this year's distribution of grades and could be different the next year
- Look at impact on each cell to see what has a greater impact on working with certain students that currently reside in one or another category
- Other states, in looking at scores, have considered what makes sense for their state as they look at both scenarios
- SEM based on conditional measurement and the estimated amount of error to be expected for a particular scale score

- Looking for a continuum as students advance through the grades, and in SEM the concept and philosophy behind the application differs between smoothing and SEM
  - One looks at grades and one makes some allowance for the student in terms of policy
  - Used to limit the range for smoothing process and eliminate false positives and false negatives
  - “Approaches” or “falls far below” are more likely to be closest to the movement
  - Not as many students at the bottom of the range so the biggest jump will not be seen at this level

#### Math:

- Dip in grade 8 did not make sense so cut score was moved to almost 64% of raw score across the grades and more in line with what happened in grades 3 and 5
- Rationale for recommendations is included in the memo to the State Board of Education from the Mathematics Table Leaders
  - 5<sup>th</sup> grade raw score increased by 2 items and 8<sup>th</sup> grade by 4 items
  - Are we adjusting for the instructional implication of smoothing?
  - Looked at a logical progression of percentages in each area but in 8<sup>th</sup> grade there was a very different percentage level which was moved to more accurately measure how kids are progressing from the 8<sup>th</sup> grade to the 10<sup>th</sup> grade
- Three types of Standard Error that are computed:
  - Based on Bookmark procedure
  - Test Standard Error of Measurement
  - Combination of the two types

#### Writing:

- Writing teachers graded papers
- Had four rounds to show writing folks what happened last year
- After the fifth round the final data results were submitted
- Tables chose not to change cut scores as per memo attached
- Why are these not consistent with the other areas?
  - The committee felt that the final round of scores based on a holistic view accurately reflect the judgment of the Arizona educators who comprised the larger writing committee
  - Decided not to smooth as they see a different concept.

Ms. Irene Hunting, Director of Test Administration and Contracts, Accountability Division, Arizona Department of Education stated there was some concern that the teachers represented only 25% of all educators and community members involved in standards setting and felt they were not being true to the 75% who were not there to move cut scores. Table leaders saw this as part of their task.

- Another table leader noted that they talked about smoothing scores and had difficulty with lowering the bar or changing the standards and they couldn't back off from those standards
- Ms. Marian Carol, a parent representative and table leader, saw variation in 5<sup>th</sup> & 8<sup>th</sup> grades but took into account:
  - things happening at the 5<sup>th</sup> and 8<sup>th</sup> grade ages
  - respect for the professionals who carefully set these values originally
  - nice to make it pretty but not representative of the process
- NAAAC agreed to bring all educators' recommendations to the State Board

- There are different forms of writing between all grades
- Dr. D'Agostino noted the following:
  - It is hard to tell what effect smoothing would have had; could lower or raise scores
  - NAAAC did not make further recommendations as the educators' recommendations came at the end of NAAAC's meeting and there was limited time for further consideration
  - NAAAC recommended smoothing in reading and math
  - Have a vertical scale to consider and if smoothing isn't done might get some strange results across grades
  - Performance levels may vary from grade level to grade level
  - Smoothing allows you to borrow information from the ensemble of data that's present
  - Standards setting committee only looked at grades 3, 5, and 8 and have to interpolate standards in the non-standards-setting grades via smoothing
    - This would indicate that smoothing should be done in all grades to be uniform
- How can one portion of the process decide that smoothing is not relevant when the rest of the educators say it is
- Don't know what the result would have been if smoothing had occurred. Don't think NAAAC would have suggested smoothing in writing. NAAAC's posture is that a sound process was used to review test content and set performance standards. Position of NAAAC is to go along with recommendations. Process should be valid.
- Writing results show a drop from 8<sup>th</sup> grade to 10<sup>th</sup> grade and high school students have multiple opportunities
- Could seem that since only 25% of educators were represented there was not consensus in the 25% and therefore the best option is not to smooth versus having 100% of the 25% present feeling strongly about not smoothing.
- Was there lack of consensus in the 25% or did they clearly feel that smoothing was inappropriate?
  - Not consensus in the room to not smooth but some would have been happy to smooth
  - Assumption that smoothing would lower the bar and some of the conversation was about raising the bar
  - Consideration was given to smoothing between "approaches" and "meets" for 8<sup>th</sup> grade and between "meets" and "exceeds" for the 5<sup>th</sup> grade
    - Issue in 5<sup>th</sup> grade is no longer apparent
    - Issue in 8<sup>th</sup> grade is still there and the discussion was not to lower the cut score, but to raise it for those who wanted to smooth
- Different method than reading or math; not exact comparison; may be the result of a poor writing prompt
- SBE can accept recommendations or not, but its interest to be consistent is understood
- NAAAC cannot shadow every committee, so it takes on faith that every process was followed
- Not technically unsound for the SBE and other agencies to deviate from the standards setting committees' recommendations.

- Technical review committee is saying process is sound and based on what they know; they support the standards setting committee's recommendations.
- SBE could ask what the distributions would look like assuming smoothing occurred- some kind of smoothing process. Look at some statistically-driven processes.
- Ms. Francine Prather, a participant in the high school writing process, addressed the State Board:
  - Very comfortable with the integrity of the process
  - Highly impressed by the educators' professional knowledge and very high standards that did not seem to be unreasonable for today's 10<sup>th</sup> graders
  - Students had less opportunity to practice persuasion until essential standards were adopted; they probably had less opportunity
  - 10<sup>th</sup> grade prompt was a problem this year
  - Pleasantly surprised at the quality of writing, given what was expected based on the prompt
  - Highly confident in cut scores that created a fair and extremely accurate measurement of the standards; quantified the creative process based on what is being taught and what students should know
  - Committee's concern was to see more exceeding the standards; does not mean that we don't have excellent writers; different than what goes on in classrooms on a regular basis
  - Test is doing exactly what it is supposed to do: improve students' writing

Further discussion:

- The assessments were made based on the prompts that were given but Ms. Prather did not favor the SEM related to the measurement
- Ms. Kimberlyn Hicks, a participant in the 8<sup>th</sup> grade reading process addressed the State Board:
  - She took the AIMS test; could see the numerous strategies students are asked to implement
- This impacted where bookmarks were put
  - Were not given negative or positive feedback by the testing company on the numbers they came up with
  - Decisions were based on where they would place the bookmarks on standards
  - Curriculum being taught has more and more standards being applied
  - Might have hit some standards in her early teaching career but is able to hit more of them now

Further discussion:

- In looking at the best fitting line the SBE should also look at it in reading and math, but don't think it would have an impact
- SBE should consider all the additional information
- Cut score in writing was higher in grade 5 in "approaches" than in grade 8
- Would expect cut scores to increase as they move along the grades
- If grade 8 raw score is changed from 10 to 11 the "approaches" cut score would be 398 and the increase would be seen again as we go through 8<sup>th</sup> grade
- Right now the score is going down and needs to go up

- All prompts field tested a year ago and high school prompt performed very similar to other prompts used in the past

*Break at 8:15 Pm. Reconvened at 8:30PM*

Further discussion:

- Based on knowledge some smoothing will occur in writing as does not make sense on scale

Dr. Karla Egan recommended that some smoothing occur in writing as it doesn't make sense right now on a vertical scale.

Ms. Lissa Watts, a participant in the 8<sup>th</sup> grade Reading committee and table leader addressed the State Board:

- Discrepancy between '04 and '05 but she feels better about it now as it is now more closely aligned to the standards
- Students are better prepared and more familiar with AIMS
- Teachers are more familiar with the standards than when they were first introduced
- The test is more realistic as to how students are being prepared.

Discussion as to the direction the State Board of Education should go:

Superintendent Horne noted that it is not technically wrong for the State Board of Education to make findings different from the standards setting committee due to policy considerations the Board must consider.

Mr. Horne's first recommendation based on where he differs from the standards setting committees:

- For grades 3-8, accept all recommendations except to recommend some smoothing in writing as recommended by CTB
- For high school, based on policy considerations, not make recommended reductions in reading and math but keep the percent necessary to attain proficiency on high school tests constant at 72% for reading and 71% for math
  - scores will show that students are doing a lot better and is partially masked when the cut score is reduced
  - The '06 class has 63,500 students and without high stakes test the number of students expected to graduate could be calculated based on looking at the conditional graduation rate:
    - Number of students who graduated in 4 years of high school considering a 3 year average to eliminate spikes, is 72% or 45,720 students
  - With standards recommended by the standards setting committee (in reading for the reduction of the cut score to 59% correct) 55,000 would pass reading
  - In math if the recommended 60% rate is used, 46,400 would pass. If last year's cut scores were used with one SEM (and NAAAC does not agree with this), 52,000 reading students would pass and 43,500 math students would pass.
  - Without changing the cut scores and based on the fact that students are studying and are tutored, almost 100% of those students would graduate passing the math test.
  - The numbers of students that would graduate passing all three tests is not available. Estimated on October scores where 7% of those students did not pass the math test the total number graduating may be short by 7%.

- There are still two more opportunities for students to take the tests and the legislature may pass a bill providing that grades can equal up to 25% of the student's points on the AIMS test.
- Almost 100% of students expected to pass can be achieved without changing the cut scores and if cut scores are maintained it provides more credibility for the system as a whole.
- Clarification regarding the 63,500 students in the '06 class and whether it is reasonable to use last year's cut scores:
  - Superintendent Horne noted that NAAAC advised not to use last year's cut scores but if the 3-8 recommendations are taken then the high school recommendations should also be taken. Therefore, NAAAC does not agree with the Superintendent's recommendations.
  - If the Superintendent's recommendation is chosen it would include using one SEM

Mr. Horne's second recommendation:

- Related to exceeding the standards on the high school writing test, recommendation is to increase so 8% of students would exceed, which was 4.5% previously with the following rationale:
  - Reading exceeds 8%
  - Math is substantially higher than 8%
  - Last year the exceeds numbers for writing was 2% in high school
  - Writing is a more subjective grading process and is already subject to some skepticism by the public as to the reliability of the grading
  - If they don't succeed it undermines credibility

Mr. Bruce Randall, Scientist, CTB- McGraw/Hill, reported that the Fall and Spring '04 numbers of students that passed each of the tests and then those that had passed all three tests were used with the Spring '05 number who passed, to calculate approximately how many had passed up to and including Spring '05. Approximately 50%-55% were passing all three after spring '05 administration (class of '06). Forty-five (45%) had not passed all three content areas. With additional tries, more of the 45% will pass or drop out in the senior year. Using last year's cut score for passing math comes to 70%, meaning that all kids that would ordinarily drop out may pass.

Dr. Pedicone noted the following:

- What is the prediction of how many students in the class of '06 at this point would have passed all three AIMS tests?
- Those kids who won't pass can't be discarded
- Concerned that the AIMS test should be a test that all kids can achieve
- Understands the political impact of keeping the cut scores the same as last year
- The whole smoothing process comes down to subjective decisions, which is unnerving in a way
- If a standard error of measurement is adopted what is the impact on kids that were having trouble?
- Is there a better way to look at how to treat these kids?
- Need to understand a process that has consistency over time, based on long range view

Superintendent Horne responded:

- The Department is working hard to help the 30% of kids that are dropping out
- When evaluating the impact of lowering a cut score, need to see how many are going to pass and look at how many more would graduate
- Not serving as a screen for reaching proficiency
- Obligated to say that if the cut score is kept kids are still doing a lot better

Ms. Mendoza noted that some consistency needs to be seen as they look at scores across the board.

Dr. Egan pointed out that the scale score is comparable from year to year and is not going to vary dramatically but will vary slightly.

Ms. Mendoza stated she supports high expectations and accountability but how can they go from 5<sup>th</sup> and 8<sup>th</sup> meeting the standards at 61% and 63% and 53% and 59% and then need 70's in high school?

Mr. Randall noted that the tests are going to differ in difficulty across years so that is why the item response theory scale is used. Percent correct could assume that the same amount of achievement could be the same each year. The level of difficulty differs from year to year with a difference in required percentages so it likely did not indicate the same degree of educational attainment.

Dr. D'Agostino explained that this year is different and will not be equating '04 and '05 but will be starting afresh. If equating and have comparable tests from year to year, could see similarities, but this is new and will show differences from past scores.

Mr. Randall used a football analogy to demonstrate:

- Selecting linemen looking for strength you may give weight tests
- Give 5 attempts and one gets 4 right and one gets 3 right
- Take the attempts and the player's strength on the weight scale and the winner is the one who moves more actual weight
- Need to focus on the scale score as the measure of achievement and not the percentage

Dr. D'Agostino clarified that the Superintendent was 98% correct. The NAAAC's position is to recommend the cut scores of the standard setting committees. If the Board deviates from the standard setting committee's recommendations by using another systematic procedure, i.e. the Superintendent's recommendation, it would not be technically unfeasible to use a SEM.

NAAAC did not make a statement on whether or not to use a SEM.

Dr. Nicodemus noted:

- The key would be that the best measure and appropriate cut score will have been determined in order to evaluate who "wins" as in the previous analogy
- Regarding whether AIMS represents the skill level of students is still problematic
- Would anticipate that teachers started from that point
- Subjective piece becomes the impact piece
- Philosophically why would I choose a SEM or smoothing or any other method to assist scores?
- Keep trying to make AIMS better and now it has lost some of its strength
- Difficulty is the perception as to whether the standards have been lowered

Ms. Mendoza:

- Need to honor what the teachers are saying, then know how to support that
- Trying to use AIMS as measurement and knowing how to justify changing what teachers are recommending



Mr. Randall summarized what needs to be done now:

- Take an overview of things, using all data
- Make sure the test provides meaningful information
- See if performance levels are making sense
- If expectation really increases from 7 to 8 then it is meaningful

Dr. Pedicone:

- Don't make changes arbitrarily
- Is this test doing what it is intended to do?
- The way the test is being used changes the landscape
- All decisions have to make good logic.

Dr. Diethelm noted the simplicity of the discussions to this point:

- Groups have come up with some recommendations that they have clearly testified that they believe in
- This is a new test and comparing to the past is on shaky ground
- Superintendent Horne pointed out one technical issue regarding the writing scale
- Policy issues regarding whether the Board follows the teachers' recommendations:
  - SBE will have to be willing to say the test is different
  - Teachers are doing a better job of teaching and aligning to the standards
  - Testing is better regarding testing the standards
  - Students are learning better how to adapt their learning style and succeed better on the test
  - Therefore, these results are better
- The public will say the test was dumbed down
- The SBE can't make decisions based on whether it will be accepted by the public but needs to make the right decision and be ready to defend it

Ms. Hilde noted some details to be considered:

- Can CTB correct cut scores that would be expected to rise
  - After smoothing all others show the SBE what that would look like
- Have read all of the 8<sup>th</sup> grade writing which was done with integrity
- The SBE's level of integrity calls for the Board to look at how scores would look like with that comparison

Dr. Pedicone:

- When looking at smoothing consider "falls far below" as well as "meets"

Ms. Hilde:

- Important not to mention again what legislation may occur
- The legislation that is out there is a two-year window and does not impact what the SBE does tonight
- The integrity and sanctity of the past three days never considered possible legislation

Dr. Diethelm:

- Asked for other members' opinion relative to the perception of dumbing down of test, etc.

Mr. Ary:

- Still have concerns regarding the group of students that will appear in "falls far below"
- The only way to elevate the total student body is to elevate those who are falling below

- How different will we be, regarding concern about our kids, and what per cent of raising the bar will be accomplished with fewer dropouts and fewer students in the “falls far below” category
- Our economic system is based on a 7% unemployment rate in order to have a comfortable economy in this country
- Does that also mean that we have to have a 10% dropout?

Dr. Pedcone:

- Concerned about political implications of opinions regarding dumbing down the test
- Would rather spend energy defending why the SBE does what it does

Ms. Hilde:

- Process was not to give the SBE a second option but was rather a process to gather data
- Not comfortable using last year’s number that is not based in this year’s reality
- Process validated:
  - The test
  - Increased student learning
  - Teachers’ ability to teach around the new standards
  - Professional development results
- Hesitant to move numbers and will defend that the SBE was not dumbing down the test.
- Committees were sure the scores were placed correctly

Dr. Diethelm:

- Concern that perception is that students in schools are all of a sudden wondrously achieving and no more extra time needs to be spent to improve the quality of education in Arizona, which would be a drastic unintended consequence

Ms. Basha:

- Have seen a developmental process over the years
  - Test being developed and refined
  - Culture change in schools and with teachers
  - Process will be evolving more, will be ongoing
  - Have to refine and be able to address kids’ needs
  - SBE has confidence in what is happening in the field
  - SBE will stand on the progress to date and will not give up on any students
  - Continue to build on what is happening today

Dr. Pedicone:

- Hear assumption built in that unless pressure is continued on schools, teachers, and other that they will abandon this agenda
- Not hearing this from teachers and administrators.
- People are passionately into the fact that they are getting better at what they do
- Have to have faith in the field that they are doing a better job and are not going to stop

Dr. Diethelm:

- Not advocating pressure
- Advocating more support at all levels

Ms. Owen:

- More compelled to honor integrity of teachers who set cut scores
- Understand public perception
- Needs to be prepared to:

- Talk about what teachers say happens
- Re-double efforts to focus on better professional development activities
- Targeted strategies for teachers in classrooms
- Mapping curriculum
- Improving teachers' ability to affect change
- As continued improvement is seen, should be honest about what happens and stay the course that as been successful in many ways
- ADE, schools, teachers are committed
- Phenomenal job has been done to date given the high standards and low budgetary allotments from legislature
- Integrity has been built into the system by the Superintendent and don't want to see him back off

Dr. Nicodemus:

- Her perspective has been changed somewhat
- Still has issues with single high stakes test
- Trying to hold schools accountable
- Educators have said don't diminish AIMS
- Based on information received as new Board member, comfortable with AIMS and the intent of the work taking place around the strands, etc.
- Concerned regarding the impact on the class of '06 and whether this is a fair expectation
- Wants to do this right, trusting the experts, and know what AIMS is intended to measure
- Wants to accept the recommendations but can't be absent in the discussion about whether the '06 class been given a fair chance
- There is a difference between smoothing and standard error
- Have a better understanding but not ready to make a decision, but something needs to be done

Ms. Mendoza:

- Never going to give up on any kid although some are more challenging
- Accountability of AIMS is great but if student is not meeting standards at 8<sup>th</sup> grade, we cannot hold them responsible in high school for a system that has not helped them to move from that position
- SEM makes more sense than smoothing
- Believe AIMS is a good measure of standards and is a reasonable expectation
- Used and applied will continue to improve education

Ms. Owen:

- Can't ignore the Morrison Report regarding ethnicity, demographics and poverty level of students where there has been significant progress but needs to go farther

Ms. Kramer:

- Never more proud to be a teacher than after today
- Watched smoothing and noted integrity
- Teachers knew what they were supposed to teach not worrying about public opinion
- Feel happy about what happened today

Dr. Diethelm:

- Asked experts to look at smoothing in writing

- Need recommendation on the Superintendent's request for how to increase the "exceeds" category in high school writing
- Two options:
  - Superintendent Horne's suggestion regarding high school scores
  - Accept the teams' recommendations with the exception of writing as noted above

Ms. Owen:

- Need to see numbers in writing for consideration that have been referred to by Superintendent Horne

Ms. Hilde:

- Would like to see additional impact data available as it is not disaggregated

Dr. Egan:

- Writing will be smoothed and brought tomorrow
- Thanked the SBE for its involvement
- Appreciated conversations and input

Superintendent Horne:

- Conveyed to ADE staff the compliments from SBE members regarding their superb job they did the past three days
- Expressed appreciation to CTB and complimented a job well done

ADJOURN

Motion to adjourn by Dr. Nicodemus and seconded by Ms. Mendoza. *Motion passes*  
*Meeting adjourned at 9:50PM.*

**\*\*\*NOTE: Additions to the agenda are bolded and underlined.**